NPS-53Jy74121

NAVAL POSTGRADUATE SCHOOL

Monterey, California

ON PREDICTION INTERVALS FROM NORMAL SAMPLES

Toke Jayachandran

December 1974

Final Report for Period January 1974-December 1974

Approved for public release; distribution unlimited

red for:
of Naval Research, Arlington, Virginia 22217

NAVAL POSTGRADUATE SCHOOL
Monterey, California

Rear Admiral Isham Linder
Superintendent

Jack R. Borsting
Provost

The work reported herein was supported in part by the Foundation Research Program of the Naval Postgraduate School with funds provided by the Chief of Naval Research.

This report was prepared by:

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: NPS-53Jy74121	2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle): On Prediction Intervals From Normal Samples	5. TYPE OF REPORT & PERIOD COVERED: Technical Report 1/74 - 12/74
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): Toke Jayachandran	8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 61151N; RR000-01-50; PO 3-0045
11. CONTROLLING OFFICE NAME AND ADDRESS: Chief of Naval Research, Arlington, Virginia 22217	12. REPORT DATE: 12/74
	13. NUMBER OF PAGES
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report): UNCLASSIFIED
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; Distribution Unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

Approved for public release; Distribution unlimited.

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Prediction intervals, random samples, correlated samples, normal distribution

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

It is shown that the prediction intervals derived by Hahn [2,3,4] and Hickman [5] for the mean, standard deviation as well as all the observations of a future random sample based on an earlier informative random sample are valid even when the sample observations are correlated and have a specified correlation structure such as interclass correlation.

SUMMARY

It is shown that the prediction intervals derived by Hahn [2,3,4] and Hickman [5] for the mean, standard deviation as well as all the observations of a future random sample based on an earlier informative random sample are valid even when the sample observations are correlated and have a specified correlation structure such as interclass correlation.

1. Introduction

Suppose that a random sample of size $N = n_0 + n_1$ is drawn from a normal distribution $N(\mu,\sigma^2)$. The subsample consisting of the first n_0 observations will be called the initial sample and the remaining n_1 observations is called the future sample. Hickman [5] considered the problem of obtaining forecast intervals for the mean $\bar{X}_N$ and variance S_N^2 of the entire sample of size N based on the mean $\bar{X}_{n_0}$ and variance $S_{n_0}^2$ of the initial sample; the end points of the forecast intervals for $\bar{X}_N$ and S_N^2 depend only on the observations of the initial sample via $\bar{X}_{n_0}$ and $S_{n_0}^2$. Hahn [2,3,4] derived prediction intervals for the mean and variance of the second sample as well as a simultaneous prediction interval to contain each of the n_1 observations of the second sample. He also considered the problem of constructing simultaneous prediction intervals for the variances of each of k additional random samples of size n_1 based on an informative random sample of size n_0. In this paper we show that the results of Hickman and Hahn for random samples are valid even where the sample observations are correlated and have a special correlation structure such as interclass correlation. As an example, samples with interclass correlation occur in the study of random effects models in Analysis of Variance.

2. Notation And Basic Results

Let X_{ij}, $j = 1,2,\dots,n_i$, $i = 0,1,2,\dots,k$ be $(k + 1)$ sets of random samples of size n_i from a normal distribution $N(\mu,\sigma^2)$ and let $N = \sum_{i=0}^{k} n_i$. The means and variances of the $(k + 1)$ sets of random samples and the pooled sample of N observations are

$$\bar{X}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij} \qquad S_i^2 = \frac{1}{n_i} \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2 \qquad i = 0,1,\ldots,k$$

$$\bar{X} = \frac{1}{N} \sum_{i=0}^{k} \sum_{j=1}^{n_i} X_{ij} \qquad S^2 = \frac{1}{N} \sum_{i=0}^{k} \sum_{j=1}^{n_i} (X_{ij} - \bar{X})^2$$

It is convenient to use matrix notation and express the sample variances as quadratic forms, in deriving the results. Let the symbols $\underline{I}$ and $\underline{E}$ represent the identity matrix and a matrix all of whose elements are unity respectively. Also, define the vector $\underline{X}$ by

$$\underline{X} = (X_{01}, X_{02}, \ldots, X_{0n_0}, X_{10}, X_{12}, \ldots, X_{1n_1}, \ldots, X_{k1}, X_{k2}, \ldots, X_{kn_k})'$$

Then,

$$NS^2 = \underline{X}'\underline{B}\,\underline{X}$$

and

$$n_i S_i^2 = \underline{X}'\underline{B}_i\,\underline{X} \qquad i = 0,1,2,\ldots,k$$

where $\underline{B} = \underset{N\times N}{\underline{I}} - N^{-1} \underset{N\times N}{\underline{E}}$

and the $\underline{B}_i$ are $N\times N$ block diagonal matrices with the i^{th} diagonal block equal to

$$\underset{n_i\times n_i}{\underline{I}} - n_i^{-1} \underset{n_i\times n_i}{\underline{E}}$$

and the rest of the blocks being zero matrices. It is easily shown that the matrices $\underline{B}$ and $\underline{B}_i$ are idempotent and that $\underline{X}'\underline{B}\,\underline{X}/\sigma^2$ and $\underline{X}'B_i\,X/\sigma^2$

have chi-square distributions with $(N-1)$ and (n_i-1) degrees of freedom (see Rao [10, Chapter 3]. Further, the quadratic form $\underline{X}'\underline{B}\underline{X}$ can be partioned as

$$\underline{X}'\underline{B}\,\underline{X} = \sum_{i=0}^{k} \underline{X}'\underline{B}_i\underline{X} + \sum_{i=0}^{k} n_i\ (\bar{X}_i - \bar{X})^2$$

It follows from an application of Hogg and Craig's Theorem [6, Chap. XIII] that $\underline{X}'\underline{B}_i\underline{X}/\sigma^2 \quad i = 0,1,2,\ldots,k$ are mutually independent chi-square variates and that $\sum_{i=0}^{k} n_i\ (\bar{X}_i - \bar{X})^2 = \underline{X}'\underline{B}_{k+1}\ \underline{X}$ also has a chi-square distribution with k degrees of freedom; as a consequence the matrix $\underline{B}_{k+1}$ is also indempotent.

The construction of a simultaneous prediction interval to contain the variances of k sets of future samples is based on a statistic whose distribution is known as the studentized largest (smallest) chi-square distribution. Suppose Y is a chi-square random variable with ν_0 degrees of freedom, $Y_1,Y_2,\ldots,Y_m$ are identically distributed as chi - square with ν_1 degrees of freedom and $Y_0,Y_1,Y_2,\ldots,Y_m$ are mutually independent. Then, the distribution of

$$W_L = \frac{\min(Y_1,Y_2,\ldots,Y_m)}{Y_0}$$

and

$$W_U = \frac{\max(Y_1,Y_2,\ldots,Y_m)}{Y_0}$$

are known as studentized smallest and largest chi-square distributions respectively. These distributions depend on three parameters ν_0,ν_1 and m and Krishnaiah and Armitage [7,8] constructed tables of percentage points for the two distributions; the $100\gamma^{th}$ percentiles of the distributions will be denoted by $W_L(\nu_0,\nu_1,m,\gamma)$ and $W_U(\nu_0,\nu_1,m,\gamma)$.

The multivariate t distribution is used in deriving a simultaneous prediction interval for each observation in a future sample of size n_1 based on a prior sample of size n_0. Let $\underline{Z} = (Z_1, Z_2, \ldots, Z_p)'$ be distributed as multivariate normal with zero mean vector and covariance matrix $\underline{\Lambda}$; the diagonal elements of $\underline{\Lambda}$ are all equal to σ^2 and all the off-diagonal elements are equal to $\rho\sigma^2$. If S^2/σ^2 is a chi-square variate with ν degrees of freedom distributed independently of $\underline{Z}$, then the joint distribution of $t_1, t_2, \ldots, t_p$ where $t_i = \sqrt{\nu}\, Z_i/S$ is known as the central p-variate t distribution. Krishnaiah and Armitage [9] tabulated the values $t_p(\nu,\rho,\gamma)$ such that

$$P[t_i \le t_p(\nu,\rho,\gamma) \quad , \quad i = 1,2,\ldots,p] = \gamma$$

for various choices of p, ν and ρ. The percentage points $t(\nu,\gamma)$ and $F(\nu_1,\nu_2,\gamma)$ of the student's t distribution and the F distribution are also used in constructing some of the prediction intervals.

A Theorem due to Baldessari [1] used in extending the results of Hickman and Hahn is stated below.

Baldessari Theorem: Let $\underset{n\times 1}{\underline{X}}$ have a multivariate normal distribution $N(\underline{\mu},\underline{V})$ where $\underline{\mu} = (\mu,\mu,\ldots,\mu)'$ and $\underline{V}$ is a positive definite matrix and let $\underline{B}_0, \underline{B}_1, \ldots, \underline{B}_k$ be idempotent matrices satisfying

$$\sum_{j=0}^{k} \underline{B}_j = \underset{n\times n}{\underline{I}} - n^{-1} \underset{n\times n}{\underline{E}} .$$

A necessary and sufficient condition for $\underline{X}'\underline{B}_j\underline{X}/\alpha$ $j = 0,1,2,\ldots k$ to be independent and have chi-square distributions with degrees of freedom

$r_j = \text{rank } \underline{B}_j$ is that the covariance matrix $\underline{V}$ have the form

$$\underset{n\times n}{\underline{V}} = \frac{1}{2}(\underline{A} + \underline{A}') + \alpha(\underline{I} - \underline{E}) \tag{2.1}$$

where

$$\underline{A} = \begin{pmatrix} a_1 & a_1 & \cdots & a_1 \\ a_2 & a_2 & \cdots & a_2 \\ \vdots & & & \\ a_n & a_n & \cdots & a_n \end{pmatrix}$$

and α and a_i are positive constants. A covariance matrix with the structure defined in Baldessari Theorem occurs in the study of the variance component model

$$Y_{ij} = \mu + a_i + e_{ij} \qquad j = 1,2,\ldots,n \; ; \; i = 1,2,\ldots,k$$

where μ is a constant, a_i are i·i·d $N(0,\sigma_a^2)$, e_{ij} are i·i·d $N(0,\sigma^2)$ and α_i and e_{ij} are mutually independent. In this case, it can be shown that $\underline{Y}_i = (Y_{i1}, Y_{i2}, \ldots, Y_{in})'$ has a multivariate normal distribution with mean $\underline{\mu} = (\mu,\mu,\ldots,\mu)'$ and covariance matrix

$$\underline{V} = \begin{pmatrix} \sigma^2+\sigma_a^2 & \sigma_a^2 & \cdots & \sigma_a^2 \\ \sigma_a^2 & \sigma^2+\sigma_a^2 & \cdots & \sigma_a^2 \\ \vdots & & & \\ \sigma_a^2 & \sigma_a^2 & \cdots & \sigma^2+\sigma_a^2 \end{pmatrix}$$

To see that $\underline{V}$ has the same form as in Baldessari Theorem let

$$a_1 = a_2 = \cdot \cdot \cdot = a_n = \sigma^2 + \sigma_a^2 \quad \text{and} \quad \alpha = \sigma^2 .$$

3. Prediction Intervals for Sample Variances

Suppose $X_{01},X_{02},\ldots,X_{0n_0}$ is an initial sample and $X_{i1},X_{i2},\ldots,X_{in_i}$, $i = 1,2,\ldots,k$ are k sets of future samples. Let $N = \sum_{i=0}^{k} n_i$ and $\underline{X} = (X_{01},\ldots,X_{0n_0},\ldots X_{k1},\ldots X_{kn_k})'$. It is assumed that $\underset{N\times N}{\underline{X}}$ is distributed as an N-variate normal with mean vector $\underline{\mu} = (\mu,\ldots,\mu)'$ and covariance matrix $\underline{V}$ as in (2.1). The problem is to construct a simultaneous prediction interval to contain each of the sample variances S_i^2, $i = 1,2,\ldots,k$. As indicated in Section 2, if S_N^2 is the variance of the pooled sample then

$$NS^2 = \sum_{i=0}^{k} n_i S_i^2 + \sum_{i=0}^{k} n_i(\overline{X}_i - \overline{X})^2$$

or equivalently, in matrix notation

$$\underline{X}'\underline{B}\,\underline{X} = \sum_{i=0}^{k} \underline{X}'\underline{B}_i\underline{X} + \underline{X}'\underline{B}_{k+1}\underline{X} \tag{3.1}$$

For random samples i.e., for $\underline{V} = \alpha\,\underline{I}$ the variables $Q_i/\alpha = \underline{X}'\underline{B}_i\underline{X}/\alpha$ $i = 0,1,2,\ldots,k$ are distributed as chi-square and Q_i/α, $i = 0,1,2,\ldots,k$ are mutually independent; by Hogg and Craig's Theorem [6] $Q_{k+1}/\alpha = \underline{X}'\underline{B}_{k+1}\underline{X}/\alpha$ is chi-square distributed implying that $\underline{B}_{k+1}$ is idempotent.

Since in equation (3.1) the matrices $\underline{B}$ and $\underline{B}_i$ satisfy (i) $B = \sum_{i=0}^{k+1} \underline{B}_i$ and (ii) the matrices $\underline{B}_i$ are idempotent by Baldessari Theorem Q_i/α $i = 0,1,2,\ldots,k+1$ are mutually independent chi-square variates even for correlated samples with covariance matrix $\underline{V}$ as in (2.1).

The prediction intervals for S_i^2, $i = 1,2,\ldots,k$ are obtained as follows. If $k = 1$, the variates Q_0/α, Q_1/α, Q_2/α are independently distributed as chi-square with n_0-1, n_1-1 and 1 degrees of freedom respectively. The ratio $(n_0 - 1)Q_1/(n_1 - 1)Q_0$ has an F-distribution. For a specified confidence coefficient $1 - \gamma$

$$P\left[F\left(n_1-1, n_0-1, \frac{\gamma}{2}\right) < \frac{(n_0-1)Q_1}{(n_1-1)Q_0} < F\left(n_1-1, n_0-1, 1 - \frac{\gamma}{2}\right)\right] = 1 - \gamma$$

and thus

$$P\left[\frac{n_0(n_1-1)}{n_1(n_0-1)} S_0^2 \; F(n_1-1, n_0-1, \frac{\gamma}{2}) < S_1^2 < \frac{n_0(n_1-1)}{n_1(n_0-1)} S_0^2 \; F(n_1-1, n_0-1, 1 - \frac{\gamma}{2})\right] = 1-$$

A prediction interval for S_N^2 the variance of the pooled sample is obtained by noting that $(n_0-1)(Q_1 + Q_2)/n_1 Q_0$ has an F distribution with n_1 and $n_0 - 1$ degrees of freedom. A 100(1-γ)% prediction interval for $S_N^2 = (Q_0 + Q_1 + Q_2)/N$ is

$$P\left\{\left[\frac{n_1}{n_0-1} F(n_1, n_0-1, \frac{\gamma}{2}) + 1\right]\frac{Q_0}{N} < S_N^2 < \left[\frac{n_1}{n_0-1} F(n_1, n_0-1, 1 - \frac{\gamma}{2}) + 1\right]\frac{Q_0}{N}\right\} = 1-\gamma$$

The two prediction intervals for S_1^2 and S_N^2 are exactly the same as the ones obtained by Hahn and Hickman for random samples.

For $k > 1$, a simultaneous prediction interval for S_i^2, $i = 1,2,\ldots,k$ is derived by assuming that $n_i = n$, $i = 1,2,\ldots,k$. The variate

$$W_U = \frac{\max\left(\frac{Q_1}{\alpha}, \frac{Q_2}{\alpha}, \ldots, \frac{Q_k}{\alpha}\right)}{\frac{Q_0}{\alpha}}$$

has a studentized largest chi-square distribution with parameters n_0-1, $n-1$ and k. Hence,

$$P\left[Q_i \leq W_U(n_0-1, n-1, k, 1-\gamma)Q_0, \; i = 1,2,\ldots,k\right] = 1-\gamma$$

and

$$P\left[S_i^2 \leq \frac{n_0}{n_1} W_U(n_0-1, n-1, k, 1-\gamma)S_0^2, \; i = 1,2,\ldots,k\right] = 1 - \gamma$$

This simultaneous interval is an extension of Hahn's results. Lower bounds and two sided bounds may be obtained using the studentized smallest chi-square variate W_L .

4. Prediction Intervals For The Observations In A Future Sample.

Let $X_1, X_2, \ldots, X_n, X_{n+1}, X_{n+2}, \ldots, X_{n+m}$ be a sample of size $N = n+m$ such that $\underline{X} = (X_1, X_2, \ldots, X_n, X_{n+1}, X_{n+2}, \ldots, X_N)'$ has a multivariate normal distribution with mean $\mu = (\mu, \ldots, \mu)'$ and covariance matrix $\underline{V}$ as in (2.1). Let $(\bar{X}_1, S_1^2)$, $(\bar{X}_2, S_2^2)$ and $(\bar{X}_N, S_N^2)$ denote the mean and variance of the first n observations, the second set of m observations and the pooled sample of N observations respectively.

For $m = 1$, that is, $N = n+1$

$$(n+1)S_{n+1}^2 = nS_1^2 + \frac{n}{n+1}(X_{n+1} - \bar{X}_1)^2$$

or as quadratic forms

$$\underline{X}'\underline{B}\,\underline{X} = \underline{X}'\underline{B}_1\underline{X} + \underline{X}'\underline{B}_2\underline{X}$$

By the use of Baldessari theorem it can be concluded that

$$\frac{(n-1)\underline{X}'\underline{B}_2\underline{X}}{\underline{X}'\underline{B}_1\underline{X}} = \frac{n-1}{n+1}\frac{(X_{n+1} - \bar{X}_n)^2}{S_n^2}$$

is distributed as F with 1 and $n-1$ degrees of freedom. Thus,

$$P\left\{\bar{X}_n - \left[\frac{(n+1)S_n^2 \;\; F(1,n-1,\gamma)}{n-1}\right]^{1/2} < X_{n+1} < \bar{X}_n + \left[\frac{(n+1)S_n^2 \; F(1,n-1,\gamma)}{n-1}\right]^{1/2}\right\}$$

$$= 1 - \gamma$$

If $m > 1$ a prediction interval for $\bar{X}_2$ the mean of the second sample can be obtained following the same procedure as above, that is, starting with a partition of the sum of squares NS_N^2. The resulting interval would be exactly the same as the one obtained by Hickman for random samples.

To construct a simultaneous prediction interval for $X_{n+1}, X_{n+2}, \ldots, X_{n+m}$ let $Z_i = X_{n+i} - \bar{X}_n \quad i = 1,2,\ldots,m$. Then, $\underline{Z} = (Z_1, Z_2, \ldots, Z_m)'$ can be expressed as $\underset{m\times 1}{\underline{Z}} = \underset{m\times N}{\underline{C}'} \underset{N\times 1}{\underline{X}}$ where the matrix $\underline{C}'$ is defined by

$$\underset{m\times N}{\underline{C}'} = \left(-\frac{1}{n} \underset{m\times n}{\underline{E}} \;\vdots\; \underset{m\times m}{I} \right)$$

It follows that $\underline{Z}$ is distributed as multivariate normal with zero mean vector and covariance matrix (see appendix for computations)

$$\underset{m\times m}{\underline{C}'\underline{V}\,\underline{C}} = \alpha\left(\frac{1}{n}\underline{E} + \underline{I} \right) \cdot$$

Also, $\underline{Z} = \underline{C}'\underline{X}$ and $\dfrac{nS_1^2}{\alpha} = \dfrac{\underline{X}'\underline{B}_1\underline{X}}{\alpha}$ which is chi-square distributed are statistically independent since $\underline{C}'\underline{V}\,\underline{B} = \underline{0}$ (see appendix for computations) which is a sufficient condition for independence (see Rao [10, Chap. 3]).

Let

$$W_i = \frac{Z_i}{[\alpha(1 + \frac{1}{n})]^{1/2}} = \frac{X_{n+1} - \bar{X}_n}{[\alpha(1 + \frac{1}{n})]^{1/2}} \qquad i = 1,2,\ldots,m .$$ Then,

$W_1, W_2, \ldots, W_m$ are jointly distributed as an m-variate normal with zero mean vector and covariance matrix

$$\begin{pmatrix} 1 & \frac{1}{n+1} & \cdots & \frac{1}{n+1} \\ \frac{1}{n+1} & 1 & \cdots & \frac{1}{n+1} \\ \vdots & \vdots & & \\ \frac{1}{n+1} & \frac{1}{n+1} & \cdots & 1 \end{pmatrix}$$

Define the variates

$$t_i = \frac{W_i}{\sqrt{\frac{nS^2}{(n-1)\alpha}}} = \frac{X_{n+i} - \bar{X}_n}{\sqrt{\frac{n-1}{n+1}}\,S} \qquad i = 1,2,\ldots,m$$

The joint distribution of $t_2, \ldots, t_m$ is the central m-variate t distribution with parameters $(n-1)$ and $\rho = \frac{1}{n+1}$. If $t_m(n-1, \frac{1}{n+1}, 1 - \frac{\gamma}{2})$ is the upper $(1 - \frac{\gamma}{2})$th percentage point of the m-variate t distribution then

$$P\left[\bar{X}_n - t_m(n-1, \frac{1}{n+1}, 1 - \frac{\gamma}{2})\sqrt{\frac{n-1}{n+1}}\,S < X_{n+i} < \bar{X}_n + t_m(n-1, \frac{1}{n+1}, 1 - \frac{\gamma}{2})\sqrt{\frac{n-1}{n+1}}\,S \text{ for } i=1,2,\ldots m\right] = 1 - \gamma$$

This result is identical to the prediction interval of Hahn for random samples.

REFERENCES

[1] Baldessari, Bruno, "Analysis of Variance of Dependent Data", Statistica (Bologna) Vol.26, (1966), 895-903.

[2] Hahn, G. J., "Factors for Calculating two-sided Prediction Intervals for Sample From a Normal Population", J. Amer. Statist. Assoc., 64(1969), 878-888.

[3] Hahn, G. J., "Additional Factors for Calculating Prediction Intervals for Samples From a Normal Population", Journal of the American Statistical Association, 65(1970), 1668-1676.

[4] Hahn, G. J., "Statistical Intervals for a Normal Population - Parts I and II", Journal of Quality Technology, 2(1970), 115-125 and 195-206.

[5] Hickman, J. C., "Preliminary Regional Forecasts for the outcome of an Estimation Problem", Journal of The American Statiscal Association, 58, (1963), 1104-1112.

[6] Hogg, R. V. and Craig A. T., Introduction to Mathematical Statistics, Second Edition, The Macmillan Company, New York, 1965.

[7] Krishnaiah, P. R. and Armitage, J. V., "Tables for the Studentized Largest Chi-Square Distribution and Their Applications, ARL 64-188, Aerospace Research Laboratories, Wright-Patterson Air Force Base, Ohio (1964).

[8] Krishnaiah, P. R. and Armitage, J. V., "Distribution of the Studentized Smallest Chi-Square With Tables and Applications," ARL 64-218 Aerospace Research Laboratories, Wright-Patterson Air Force Base, Ohio (1964).

[9] Krishnaiah, P. R. and Armitage, J. V., "Tables for Multivariate t-distribution," Sankhya, Series B, 28, (1966), 31-56.

[10] Rao, C. Radhakrishna, Linear Statistical Inference and Its Applications, New York: John Wiley and Sons, Inc., (1965).

APPENDIX

Suppose $\underline{X} = (X_1, X_2, \ldots, X_n, X_{n+1}, X_{n+2}, \ldots, X_{n+m})'$ is a sample of size $N = n+m$ that is jointly distributed as multivariate normal with mean vector $\underline{\mu} = (\mu, \ldots, \mu)'$ and covariance matrix

$$\underset{N\times N}{\underline{V}} = \frac{1}{2}(\underline{A} + \underline{A}') + \alpha(\underline{I} - \underline{E})$$

where

$$\underset{N\times N}{\underline{A}} = \begin{pmatrix} a_1 & a_1 & \cdots & a_1 \\ a_2 & a_2 & \cdots & a_2 \\ \vdots & \vdots & & \vdots \\ a_N & a_N & & a_N \end{pmatrix}$$

and α and a_i are positive constants. Let $\underline{Z} = (Z_1, Z_2, \ldots, Z_m)'$ where $Z_i = X_{n+i} - \overline{X}_n$, $i = 1, 2, \ldots, m$. Then, if

$$\underset{m\times N}{\underline{C}'} = \left(-\frac{1}{n} \underset{m\times n}{\underline{E}} \;\vdots\; \underset{m\times m}{\underline{I}} \right)$$

$$\underline{Z} = \underline{C}'\underline{X}$$

The joint distribution of $Z_1, Z_2, \ldots, Z_m$ is multivariate normal with mean $\underline{C}'\underline{\mu} = \underline{0}$ and covariance matrix $\underline{C}'\underline{V}\,\underline{C} = \alpha(\frac{1}{n} \underset{m\ m}{\underline{E}} \quad \underset{m\ m}{\underline{I}})$ as shown below:

Partition the matrix $\underline{A}$ as

$$\underline{A} = \left(\begin{array}{c|c} \underset{n\times n}{\underline{A}_1} & \underset{n\times m}{\underline{A}_2} \\ \hline \underset{m\times n}{\underline{A}_3} & \underset{m\times m}{\underline{A}_4} \end{array}\right)$$

Note that in each of the matrices $\underline{A}_1, \underline{A}_2, \underline{A}_3, \underline{A}_4$ the columns are identical and all elements in each row are the same.

$$\underline{C}'\underline{V} = \left(-\frac{1}{n}\underline{E} \;\vdots\; \underline{I}\right)\left\{\frac{1}{2}\left[\left(\begin{array}{c|c} \underline{A}_1 & \underline{A}_2 \\ \hline \underline{A}_3 & \underline{A}_4 \end{array}\right) + \left(\begin{array}{c|c} \underline{A}_1' & \underline{A}_3' \\ \hline \underline{A}_2' & \underline{A}_4' \end{array}\right)\right]\right.$$

$$\left. + \quad \alpha\left[\left(\begin{array}{c|c} \underline{I} & \underline{0} \\ \hline 0 & \underline{I} \end{array}\right) + \left(\begin{array}{c|c} \underline{E}_1 & \underline{E}_2 \\ \hline \underline{E}_3 & \underline{E}_4 \end{array}\right)\right]\right\}$$

$$= \frac{1}{2}\begin{pmatrix} (-\frac{a}{n} + a_{n+1})\underset{1\times N}{\underline{E}} \\ (-\frac{a}{n} + a_{n+2})\underset{1\times N}{\underline{E}} \\ \vdots \\ (-\frac{a}{n} + a_{n+m})\underset{1\times N}{\underline{E}} \end{pmatrix} + \alpha\left(-\frac{1}{n}\underset{m\times n}{\underline{E}} \;\vdots\; \underset{m\times m}{\underline{I}}\right)$$

where $\quad a = \sum_{i=1}^{n} a_i$

Hence,

$$\underline{C}'\underline{V}\,\underline{C} = \left[\frac{1}{2}\begin{pmatrix} (-\frac{a}{n} + a_{n+1})\underline{E} \\ (-\frac{a}{n} + a_{n+2})\underline{E} \\ \vdots \\ (-\frac{a}{n} + a_{n+m})\underline{E} \end{pmatrix} + \alpha\left(-\frac{1}{n}\underline{E} \;\vdots\; \underline{I}\right)\right]\begin{pmatrix} -\frac{1}{n}\underset{n\times m}{\underline{E}} \\ \hline \underset{m\times m}{\underline{I}} \end{pmatrix}$$

$$= \alpha\left(\frac{1}{n}\underset{m\times m}{\underline{E}} + \underset{m\times m}{\underline{I}}\right)$$

Further if $S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$, then

$$nS^2 = \underline{X}'\underline{B}_1\underline{X} \quad \text{where}$$

$$\underline{B}_1 = \left(\begin{array}{c|c} \underset{n\times n}{\underline{I}} - n^{-1} \underset{n\times n}{\underline{E}} & \underset{n\times m}{\underline{0}} \\ \hline \underset{m\times n}{\underline{0}} & \underset{m\times m}{\underline{0}} \end{array}\right)$$

The quadratic form $\underline{X}'\underline{B}_1\underline{X}$ and the linear form $\underline{Z} = \underline{C}'\underline{X}$ are independent since

$$\underline{C}'\underline{V}\,\underline{B}_1 = \left\{ \frac{1}{2}\begin{pmatrix} (-\frac{a}{n} + a_{n+1})\,\underline{E} \\ (-\frac{a}{n} + a_{n+2})\,\underline{E} \\ \vdots \\ (-\frac{a}{n} + a_{n+m})\,\underline{E} \end{pmatrix} + \alpha(-\frac{1}{n}\underline{E} \vdots \underline{I}) \right\} \times \left(\begin{array}{c|c} \underset{n\times n}{\underline{I}} - n^{-1}\underline{E} & \underline{0} \\ \hline \underline{0} & \underline{0} \end{array}\right)$$

$$= \frac{1}{2}\begin{pmatrix} (-\frac{a}{n} + a_{n+1}) \underset{1\times n}{\underline{E}} \\ (-\frac{a}{n} + a_{n+2}) \underset{1\times n}{\underline{E}} \\ \vdots \\ (-\frac{a}{n} + a_{n+m}) \underset{1\times n}{\underline{E}} \end{pmatrix} - \frac{\alpha}{n} \underset{m\times n}{\underline{E}}$$

$$- \frac{1}{2n}\begin{pmatrix} (-\frac{a}{n} + a_{n+1})n \underset{1\times n}{\underline{E}} \\ (-\frac{a}{n} + a_{n+2})n \underset{1\times n}{\underline{E}} \\ \vdots \\ (-\frac{a}{n} + a_{n+m})n \underset{1\times n}{\underline{E}} \end{pmatrix} + \frac{\alpha}{n^2} \cdot n \underset{m\times n}{\underline{E}} = 0$$

DISTRIBUTION LIST

Defense Documentation Center (DDC) Cameron Station Alexandria, Virginia 22314	2
Library Naval Postgraduate School Monterey, California 93940	2
Department of Mathematics Naval Postgraduate School Monterey, California 93940	1
Professor Toke Jayachandran Code 53Jy Naval Postgraduate School Monterey, California 93940	25
Professor Glenn A. Stoops Code 53Zp Naval Postgraduate School Monterey, California 93940	1
Dean of Research Administration Code 023 Naval Postgraduate School Monterey, California 93940	2
Dean W. Max Woods Code 500/024 Naval Postgraduate School Monterey, California 93940	1